



Text Processing by Tamedia

Swiss Text

Marcel Blattner, Didier Orel

AUTOMATIC TEXT GENERATION

AUTOMATIC HEADLINE GENERATION

USER COMMENT WORKFLOW

TEXT TO IMAGE SEARCH



Der Bund

24heures

X homegate.ch



ricardo.ch

BZ BERNER ZEITUNG

Le Matin

autorickardo.ch

tutti.ch

Zürichsee-Zeitung

Schweizer Familie

Tribune de Genève

jobs.ch

ZATTOD

Doodle

Basler Zeitung

\$1B revenue

3600+ employees

30+ news outlets





Marcel Blattner
Chief Data Scientist
marcel.blattner@tamedia.ch



Didier Orel
Innovation Project Manager
didier.orel@tamedia.ch

A scenic view of a Swiss village with traditional wooden houses and a large green field in the foreground. The text is overlaid on a semi-transparent dark grey box.

**Do automatic texts
have some value
for our readers ?**

Hipster

39

Zuzwil will Sozialdetektive

Mit allen einverstanden zu sein, ist in der eigenen Gemeinde nicht immer möglich. Sie selber haben nur die Hornkuhinitiative befürwortet. Immerhin müssen Sie bei einer Vorlage vom Resultat nicht enttäuscht sein.

Die Zuzwilerinnen und Zuzwiler stimmen klar gegen die Initiative gegen fremde Richter, und zwar mit 67%. 95 sagen Ja und 194 Nein. Die SVP-Initiative wollte die Verfassung als oberste Rechtsquelle festschreiben, die im Konfliktfall dem internationalen Recht vorgeht. Die Gemeinde verhält sich gleich wie der Kanton Bern, der die Vorlage mit 65% Nein ablehnt. Die Schweiz sagt mit 66% Nein.

Zuzwil ist für die Überwachung von Sozialhilfebezügern, die unter dem Verdacht stehen, Missbrauch zu betreiben. Mit 67% befürworten die Stimmbürger eine Regelung dieser Frage. Bei 287 gültigen Stimmen unterstützen 192 die Vorlage. Die Gemeinde stimmt gleich wie der Kanton Bern (65% Ja). Die Schweizer sagen mit 65% Ja.

Ein Drittel für Hörner

Die Gemeinde Zuzwil lehnt die Hornkuhinitiative deutlich ab. Nur 36% der Stimmbürger stellten sich hinter das Volksbegehren. 103 Ja stehen 180 Nein gegenüber. In der Schweiz sind neun von zehn Kühen enthornt. Um diese Zahl zu verkleinern, verlangte der auf eidgenössischer Ebene abgelehnte Initiativtext des Bergbauern Armin Capaul, dass Tiere mit Hörnern subventioniert werden. Die Gemeinde stimmt gleich wie der Kanton, der die Vorlage nur mit 46% unterstützt.

Mit 64% liegt die Beteiligung der Zuzwiler 16 Prozentpunkte über dem kantonalen Durchschnitt (48%). Die Gemeinde hat die 97. höchste Beteiligung der Schweiz. Auf eidgenössischer Ebene beträgt die Stimmbeteiligung 48%.

Necks

xts

AUTOMATIC TEXT GENERATION - USE CASE VOTATIONS

W 190519_vote-DE-FR SHARED TEMPLATE DE_national_object_1 Shared in

Outline Show rule logic

Search for text Search Collapse All

- Die Stadt Aeugst am Albis nimmt die AHV-Steuer vorlage mit grosser Mehrheit an. 53,2% der Stimmenden haben Ja gesagt.
 - Die
 - Stadt
 - Stadt
 - Stadt
 - Gemeinde
 - Gemeinde
 - Aeugst am Albis nimmt die AHV-Steuer vorlage
 - mit grosser Mehrheit
 - mit grosser Mehrheit
 - ziemlich deutlich
 - ziemlich deutlich
 - äusserst knapp
 - äusserst knapp
 - knapp
 - knapp
 - an.
 - 53,2
 - 53,2
 - 53,2
 - 53
 - 53
 - % der Stimmenden haben Ja gesagt.
 - Die Stadt Aeugst am Albis lehnt die AHV-Steuer vorlage mit grosser Mehrheit ab. 46,8% der Stimmenden haben Nein gesagt.
 - Die
 - Stadt
 - Stadt

Insert Data Add Synonym Add Branch More

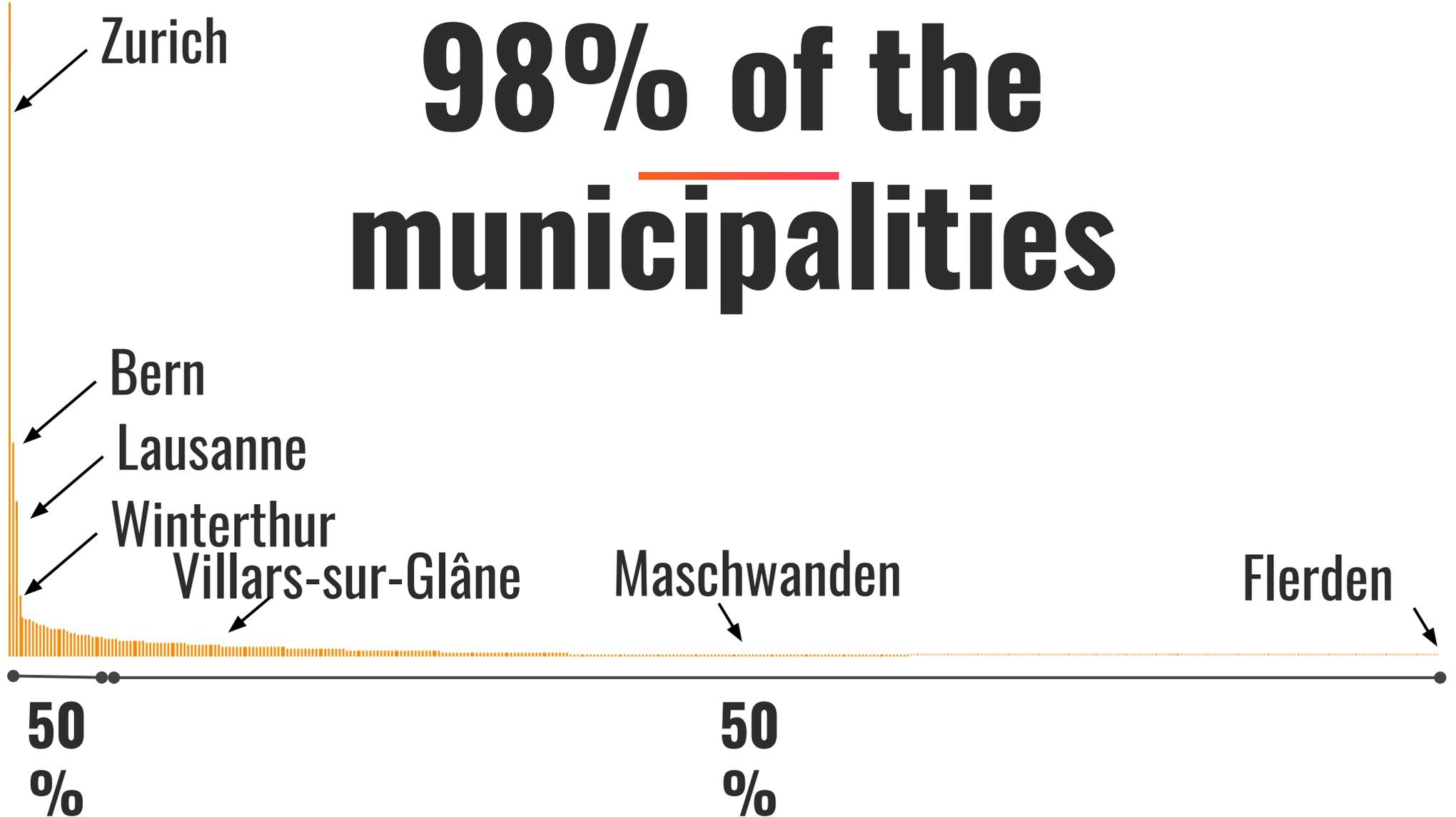
Die Stadt Aeugst am Albis nimmt die AHV-Steuer vorlage mit grosser Mehrheit an. 53,2% der Stimmenden haben Ja gesagt. Das Grossprojekt, das von der FDP, der CVP und der SP im Parlament ausgehandelt wurde, ist auf nationaler Ebene mit 54,3% angenommen worden. Das neue Gesetz soll ein attraktives Steuerklima für Unternehmen schaffen. Im Gegenzug wird eine jährliche Finanzspritze von 2 Milliarden Franken die AHV stärken.

Die Vorlage wurde von zwei Seiten bekämpft: Eine linke Allianz befürchtete riesige Steuerausfälle und kritisierte ein zu grosses Steuergeschenk an die Wirtschaft. Kreise der SVP, der Jungen SVP und der Jungen FDP, auf der andere Seite, monierten die «ungeheuerliche» Verbindung zwischen AHV und Unternehmensteuer in einem Paket.

National ist Aeugst am Albis auf dem 890. Rang jener Schweizer Gemeinden, die am deutlichsten Ja sagen zur Reform (60. Rang kantonal). Unter den grösseren Städten hat Aeugst am Albis die Vorlage am stärksten befürwortet. Das Abstimmungsresultat der Stadt fällt gleich aus wie beim Kanton, der die STAF-Vorlage mit 54,7% Ja annimmt.

Der STAF-Kompromiss entstand nach der Ablehnung der Unternehmenssteuerreform III und der Reform der Altersvorsorge vor zwei Jahren. Die Mehrheit der Bürgerinnen und Bürger von Aeugst am Albis hatte 2017 die AHV-Reform mit 51,9% angenommen. Was die USR III betrifft, sagten im Februar 2017 42,4% Ja.

98% of the municipalities



AUTOMATIC TEXT GENERATION - LOCAL SPORT USE CASE

FVBJ - Fussballverband Bern/Jura Alle Verbände

Meisterschaft - 2. Liga / Regional / Gruppe 1 - 08.06.2019 16:00 - Bodenweid, Bern - Kunstrasenfeld Bodenweid 1

 **4:4** 
(2:2)

SC Bümpliz 78 **FC Breitenrain**

18' Amadeu Prack Almeida
42' Simon Lauper
54' Mergim Shaqiri
68' Mergim Shaqiri

19' Eigentor (Christian Küng)
Christian Küng
29' Davide De Giorgi
76' Cédric Frey
80' Davide De Giorgi

Verlauf Aufstellung

90'+2'	Verwarnung Edi Cavkic (Breitenrain)
90'	Aus-/Einwechslung Bümpliz Mergim Shaqiri ersetzt durch Ardit Nasufi
88'	Verwarnung Fitim Avdijli (Bümpliz)
80'	Tor Breitenrain 4:4 Torschütze Davide De Giorgi
76'	Tor Breitenrain 4:3 Torschütze Cédric Frey
75'	Verwarnung Davide De Giorgi (Breitenrain)
68'	Tor Bümpliz 4:2 Torschütze Mergim Shaqiri
67'	Aus-/Einwechslung Bümpliz Ismael Agrebi ersetzt durch Emanuel Soares da Silva
57'	Aus-/Einwechslung Bümpliz Dean Addo Adomako ersetzt durch Fitim Avdijli
54'	Tor Bümpliz 3:2 Torschütze Mergim Shaqiri
46'	Aus-/Einwechslung Bümpliz Velid Kurtanovic ersetzt durch Jahmali Ramazani
42'	Tor Bümpliz 2:2 Torschütze Simon Lauper
29'	Tor Breitenrain 1:2 Torschütze Davide De Giorgi
20'	Aus-/Einwechslung Breitenrain Remo Bühler ersetzt durch Ali Amedoski Sadiku
19'	Eigentor Bümpliz 1:1 Eigentor Christian Küng
18'	Tor Bümpliz 1:0 Torschütze Amadeu Prack Almeida



 **SC BÜMPLIZ 78** **4** **:** **4** **FC BREITENRAIN** 
Alle Spiele

Samstag, 8. Juni 2019

2 LIGA - REGIONAL - GRUPPE 1
Punkteteilung trotz Doppelpack von Davide De Giorgi

Im Spiel des SC Bümpliz 78 gegen FC Breitenrain gab es Tore am laufenden Band. Am Ende teilten sich die beiden Kontrahenten die Punkte beim Stand von 4:4. Wer vor dem Anpfiff ein Match auf Augenhöhe erwartet hatte, fühlte sich durch den Ausgang der Partie bestätigt.

Auf die vermeintliche Siegerstrasse brachte Amadeu Prack Almeida sein Team in der 18. Minute. Bümpliz baute den Vorsprung mit der Hilfe des Gegners weiter aus. Unglücksrabe Christian Küng beförderte den Ball in der 19. Minute ins eigene Netz und erhöhte damit die Führung des Gastgebers auf 2:0. Davide De Giorgi schoss für FC Breitenrain in der 29. Minute das erste Tor. Wenige Momente vor dem Seitenwechsel war Simon Lauper mit dem 3:1 für den SC Bümpliz 78 zur Stelle (42.). Torlos verabschiedeten sich beide Mannschaften schliesslich in die Pause. Mergim Shaqiri gelang ein Doppelpack (54./68.), mit dem er das Ergebnis auf 5:1 hochschraubte. Innerhalb weniger Minuten trafen Cédric Frey (76.) und Davide De Giorgi (80.). Damit bewies FC Breitenrain nochmals die Durchschlagskraft der Offensive. Schliesslich gingen Bümpliz und FC Breitenrain mit einer Punkteteilung auseinander.

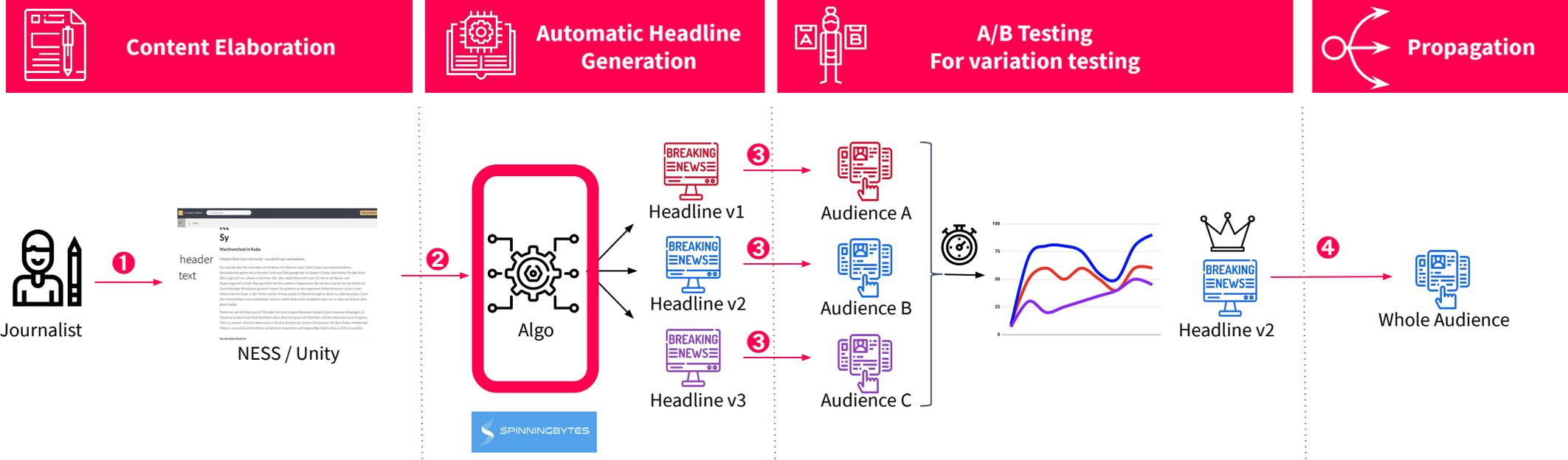
Der SC Bümpliz 78 bleibt mit diesem Remis weiterhin auf Platz neun. Bümpliz bestätigte den Gesamteindruck der Saison auch im letzten Ligaspiel. Abermals präsentierte sich die Hintermannschaft anfällig. Insgesamt kassierte der SC Bümpliz 78 50 Gegentreffer. Bümpliz beendet die Saison mit insgesamt sieben Siegen, fünf Remis und zehn Pleiten.

Zum Ende des Fussballjahres bekleidet FC Breitenrain den sechsten Tabellenplatz. Es passt ins Bild, dass die chronischen Abwehrprobleme des Gasts auch im letzten Saisonspiel zum Vorschein kamen. Unterm Strich kassierte FC Breitenrain 55 Gegentreffer. Zum Saisonende hat FC Breitenrain neun Siege, neun Niederlagen und vier Unentschieden auf dem Konto stehen. Die letzten Resultate von FC Breitenrain konnten sich sehen lassen – acht Punkte aus fünf Partien.

BZ Berner Zeitung, erstellt die Matchberichte auf Basis der offiziellen Daten des Schweizerischen Fussballverbands mit Hilfe künstlicher Intelligenz



AUTOMATIC HEADLINE GENERATION



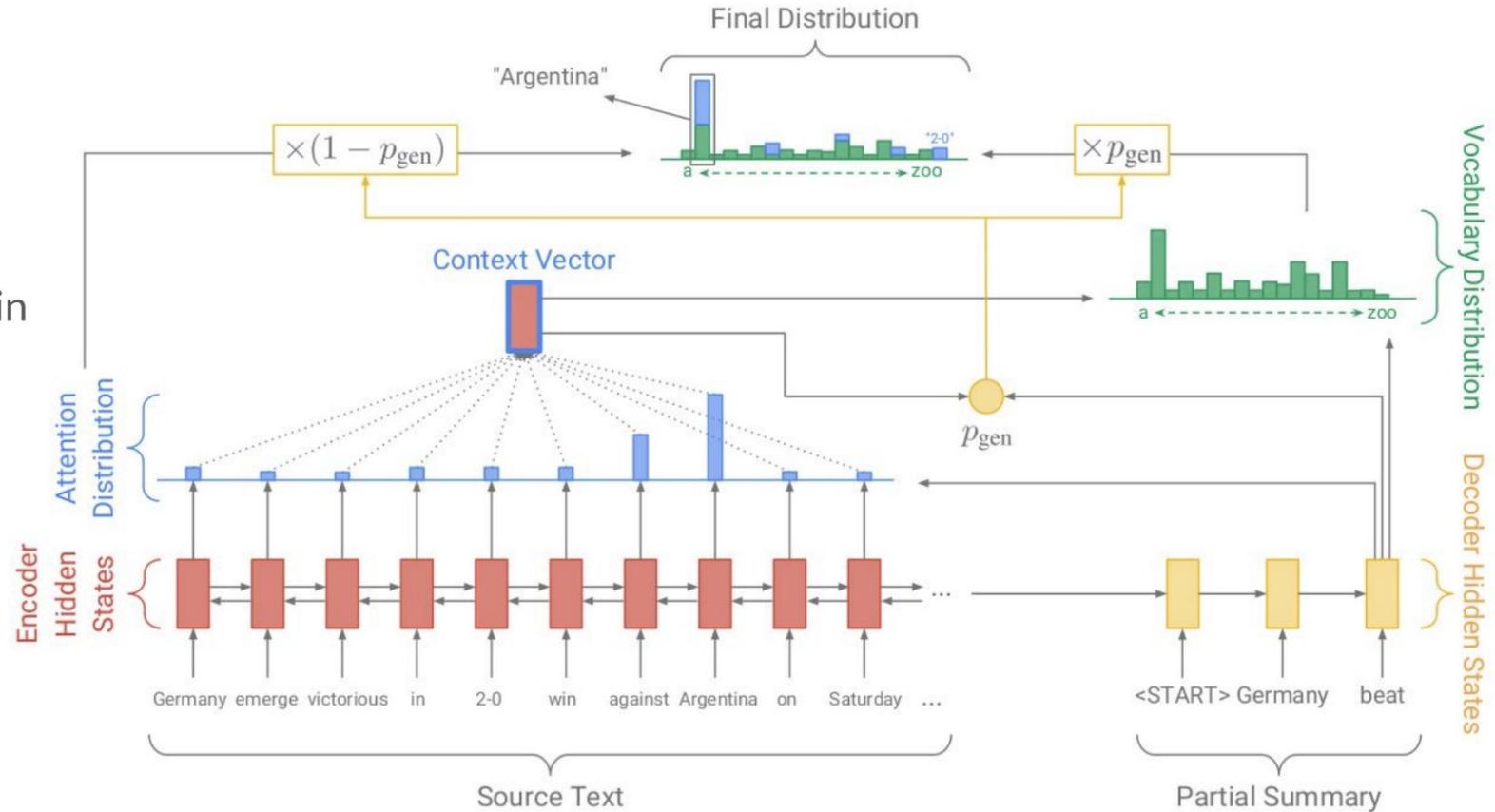
Model

Attention mechanism

- Focus on certain parts in the input article

Pointers: Two ways to emit a word

1. Generate from vocabulary
2. Copy from input





ROUGE-1	ROUGE-2	ROUGE-L
31.33	13.7	27.47

Cherry Picked

Generated	Reference
joachim low signs his contract by another two years until 2018.	germany coach joachim low extends contract until 2018 after world cup glory
jessica buchanan and poul thisted rescued in somalia	freed aid workers arrive at u.s. military base in italy

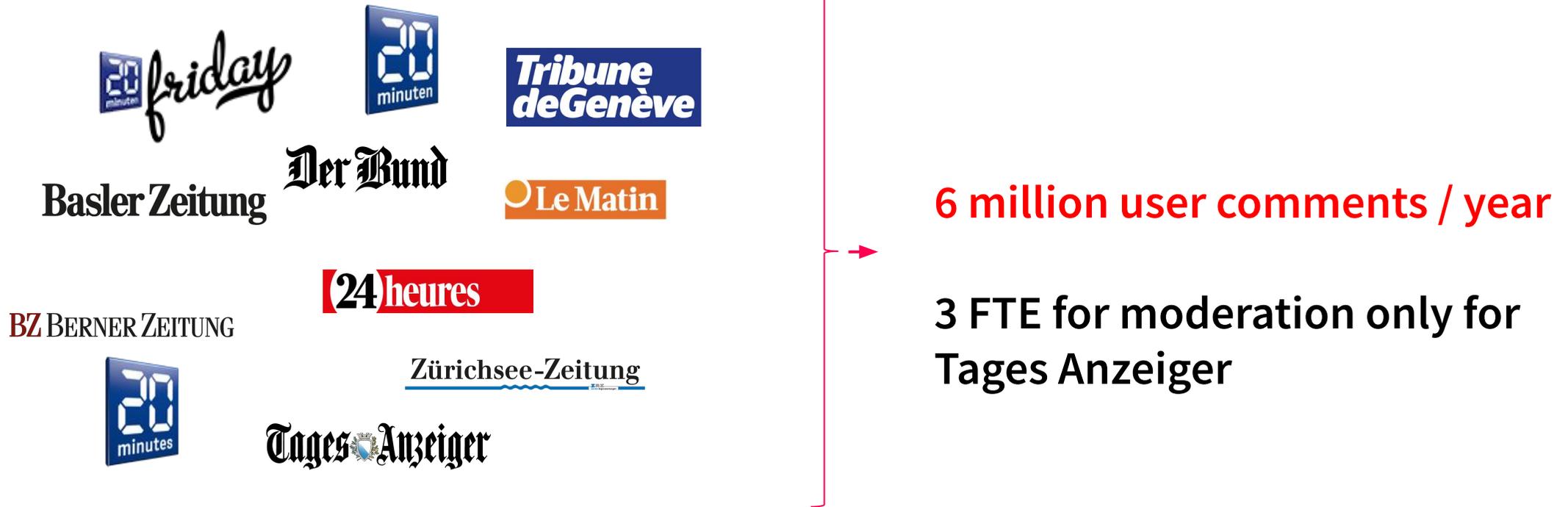


Why ?

Tages Anzeiger style is very different than the CNN/Daily mail :

- Shorter (5.6 words vs. 12.7)
- More creative (Jaccard-Similarity of 0.015 vs. 0.033)

USER COMMENT WORKFLOW



Develop comment classifier (toxic/nontoxic).

→ Goal is to reduce workload ($\approx 30\%$) for reviewers

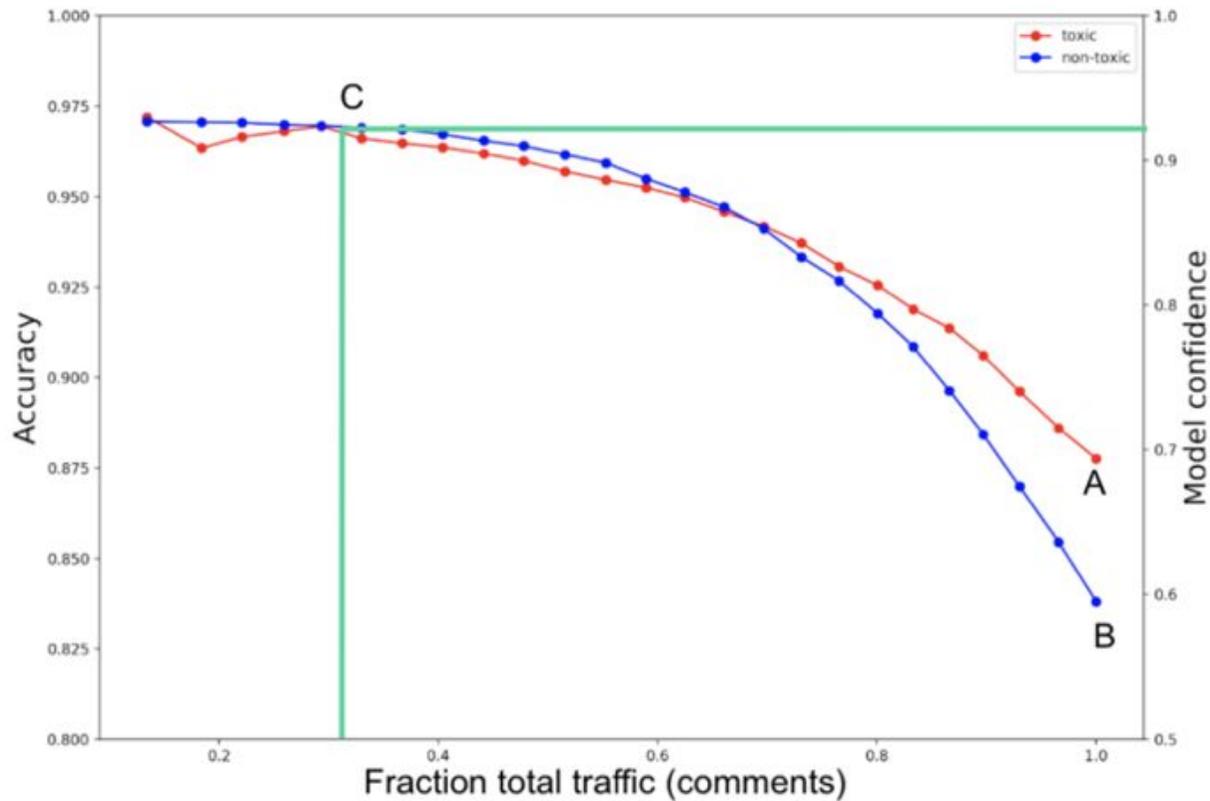
USER COMMENT WORKFLOW - ① Toxic Comment detection

Method	Accuracy (on test set)
Naive Bayes	79,2%
Logistic Regression	81,4%
CNN	86,5%

- 700K comments
- Trained on a balanced dataset - standard methodologies (80/20 splits)
- German comments only



USER COMMENT WORKFLOW - ① Toxic Comment detection



Spotted by the model but missed by the reviewer:

“Meines Wissens ist Köppels Frau keine reinrassige Schweizerin, also werden es seine Kinder auch nicht sein....

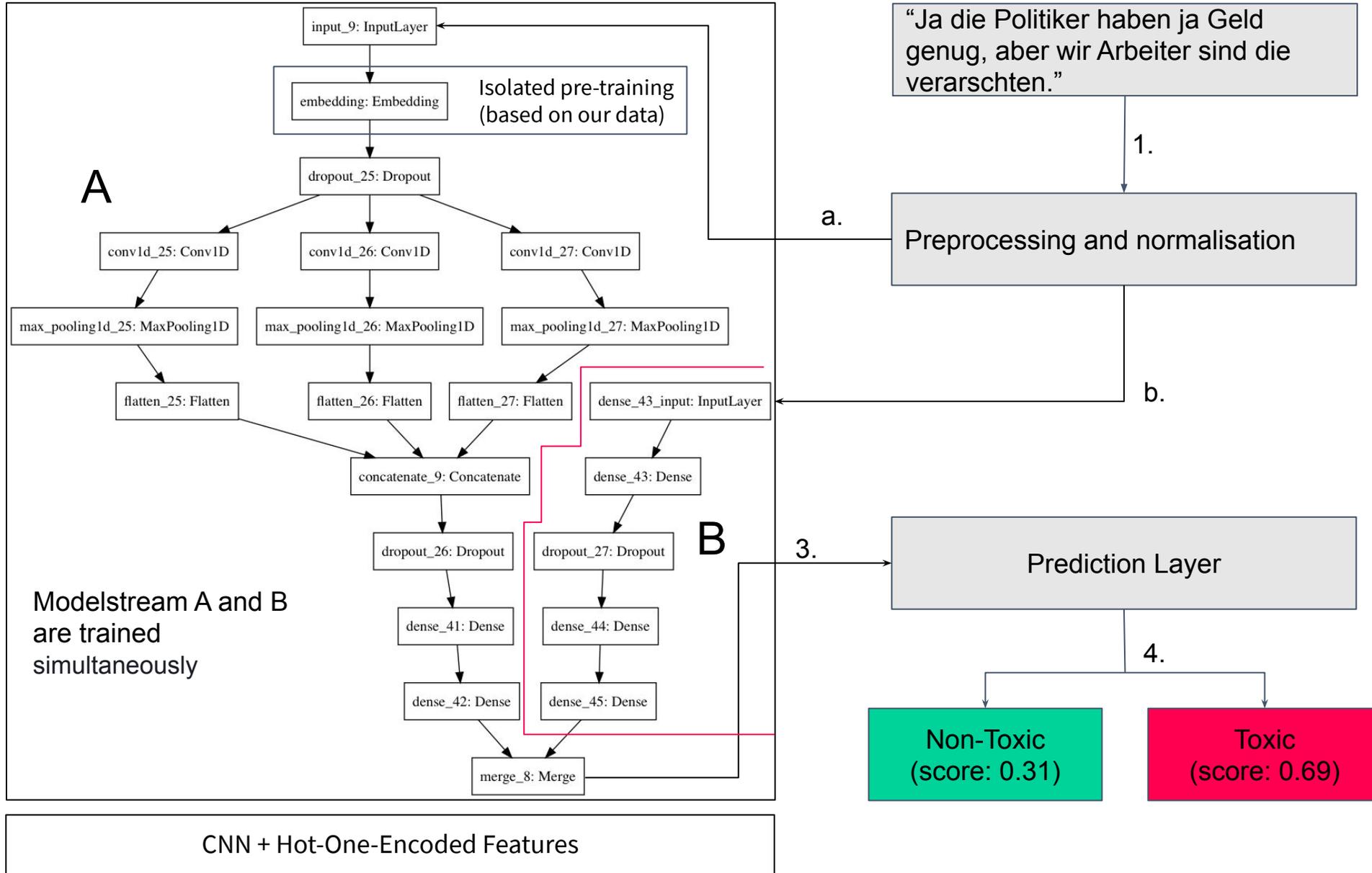
Von wem wird ein solcher Widerling gewählt?”

Toxic Score: 0.89

Limitations and challenges

- Topic drift forces to re-train the model
- Deployment
- Monitoring

USER COMMENT WORKFLOW - ① Toxic Comment detection - Model



- Can we spot some significant model bias and stereotypical associations between neutral words and female/male politicians?
- Define association strength between normalised vectors based on embedding process

Association strength:

$$as(\hat{w}, S) = \sum_{s \in S} \langle \hat{w}, \hat{s} \rangle / |S|, \quad s \in S$$

$$S_{neg} = \{wendig, brav, eitel, kokett, narzistisch, launig, feige, skrupellos \dots\}$$

$$S_{pos} = \{flessig, geschickt, aktiv, clever, elegant, gewandt, klug, qualifiziert \dots\}$$

$$|S_{pos}| = |S_{neg}| = 100$$

Normalized word vector:

$$\hat{w} = \frac{w}{\|w\|}, \quad w \in \mathbb{R}^n$$

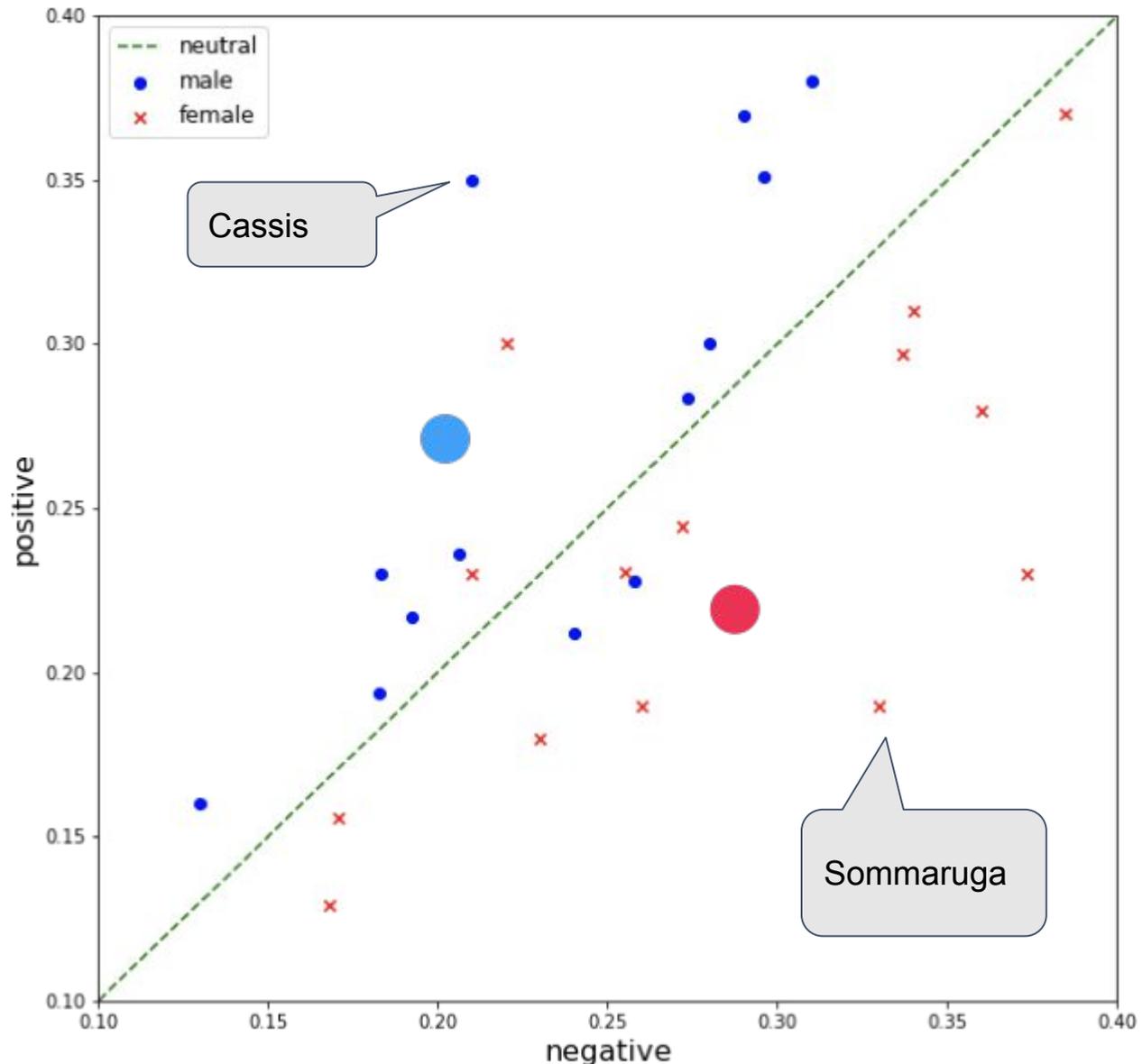
Similarity:

$$\text{sim}(\hat{w}, \hat{u}) = \langle \hat{w}, \hat{u} \rangle$$

Association strength:

$$\text{as}(\hat{w}, S) = \sum_{s \in S} \langle \hat{w}, \hat{s} \rangle / |S|, \quad s \in S$$

USER COMMENT WORKFLOW - ① Toxic Comment detection - Model Bias

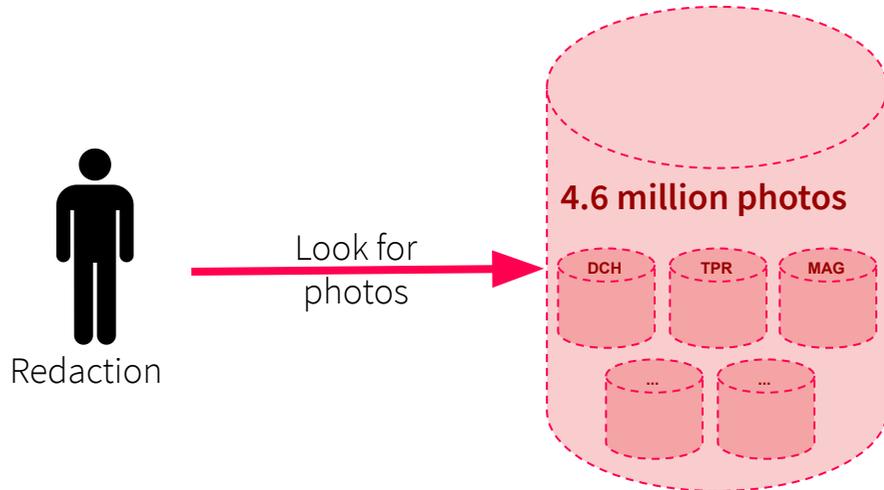


- Negative association strength vs. positive association strength (sample) of male and female politicians
- Female politicians are more likely associated with negative gender neutral words

- Average over all female politicians
- Average over all male politicians

Removing the bias results in a model accuracy decrease (10%)

TEXT TO IMAGE SEARCH



Current search

Current search tools are based on **Metadata fields**

Limitations :

- **Tagging** is **heterogeneous**
- **Impossible** to make **multilingual search**

Let's rock Machine Learning

Mix of

- Text Embedding
- Image Embedding

We partnered with



to study this question



1. Self-Supervised Learning

*“Self-supervised learning is autonomous supervised learning. It is a representation learning approach that **eliminates the prerequisite requiring humans to label data**. Self-supervised learning systems **extract and use the naturally available relevant context and embedded metadata as supervisory signals**.”*



- Images are described by noun phrases:
 - **students**
 - **their iPads**
 - **the Steve Jobs school**

Caption: "***Students*** play with ***their iPads*** at ***the Steve Jobs school*** in Sneek August 21, 2013.”

2. Multi-lingual

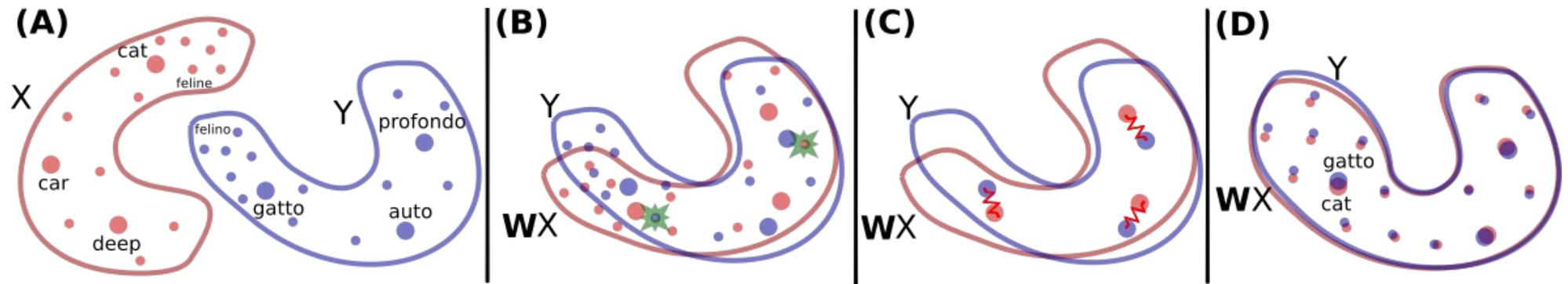


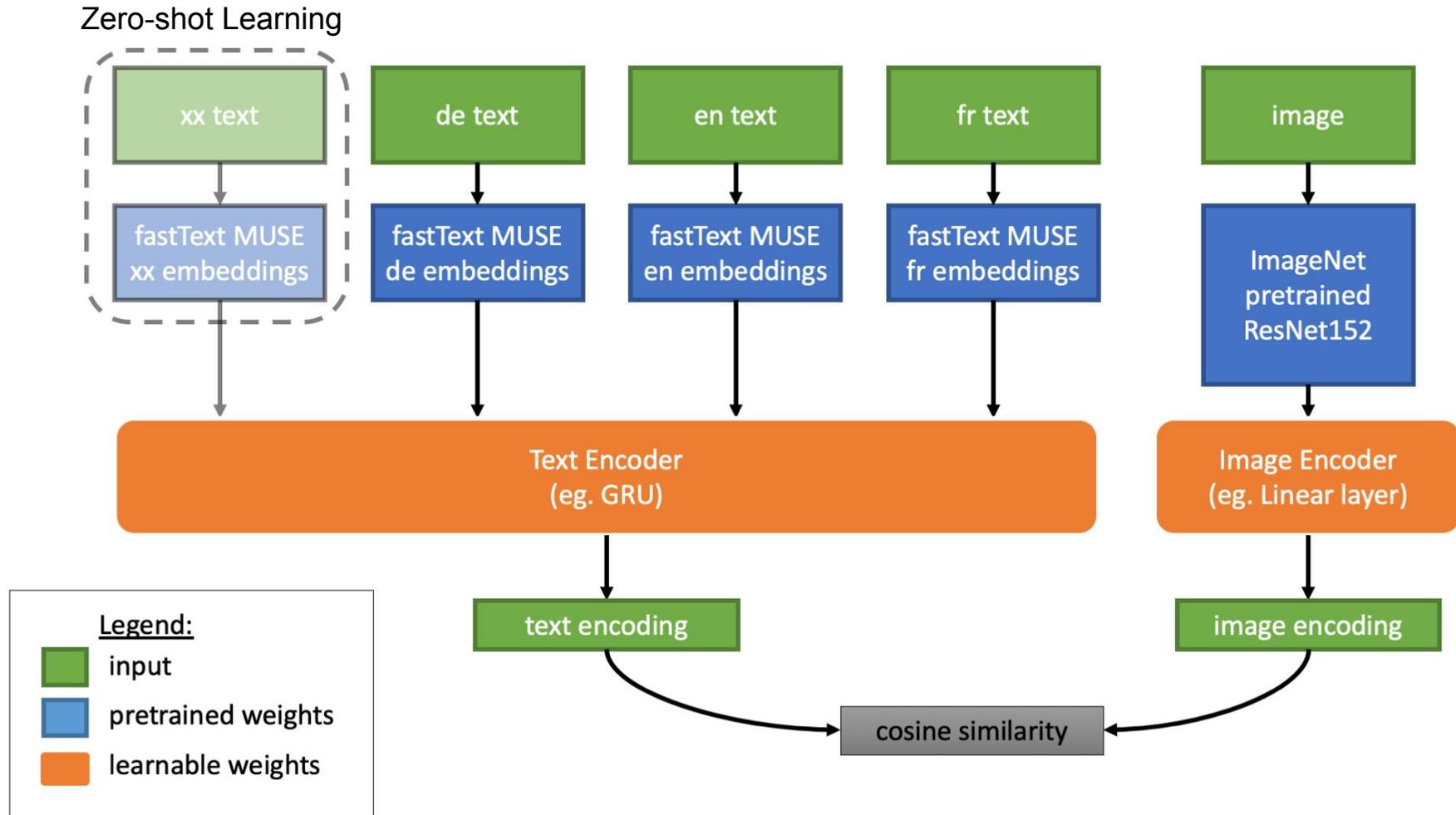
Figure from paper “Word Translation without Parallel Data”, Conneau et al., ICLR 2018
Code available at: <https://github.com/facebookresearch/MUSE>

Aligned word vectors freely available at: <https://fasttext.cc/docs/en/aligned-vectors.html>

Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion

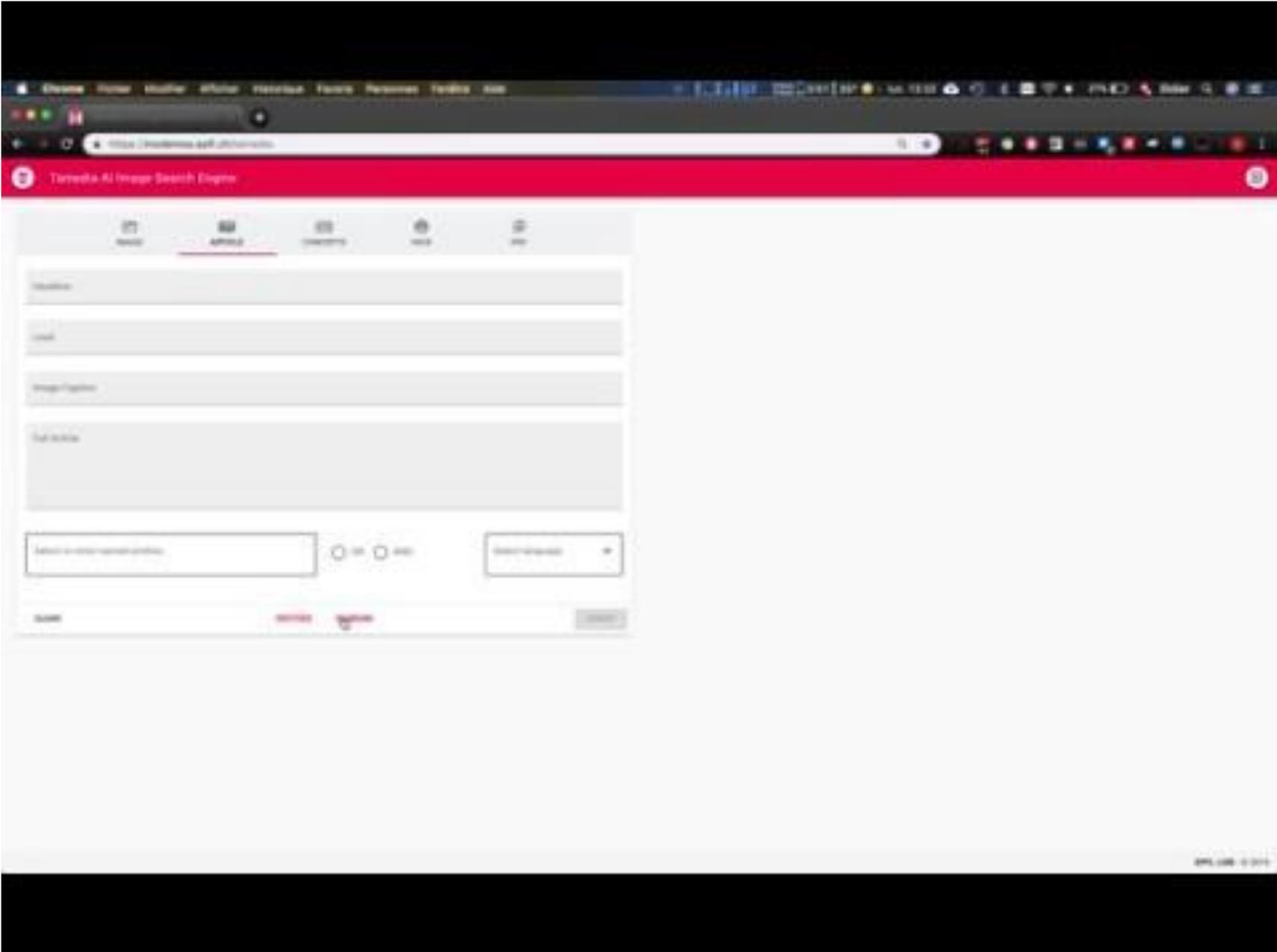
Joulin et al., EMNLP 2018

3. Map embedded noun phrases and embedded images



TEXT TO IMAGE SEARCH

Next



2